REMARKS

Applicant appreciates the Examiner's thorough consideration provided in the present application. Claims 1, 2, 6, 7, 9, 10, 14-16 and 18-20 are now present in the application. Claims 1, 6, 9, 14 and 15 have been amended. Claims 3-5 and 11-13 have been cancelled in this Reply. Claims 18-20 have been added. Claims 1 and 9 are independent. Reconsideration of this application, as amended, is respectfully requested.

Claim Rejections Under 35 U.S.C. §§ 102 and 103

Claims 1-4, 8-13 and 17 stand rejected under 35 U.S.C. § 102(b) as being anticipated by D'Hoore (U.S. Patent No. 6,085,160); claims 5 and 13 stand rejected under 35 U.S.C. § 103(a) as being unpatentable over D'Hoore in view of Burns (U.S. Patent No. 5,454,106); claims 6, 7, 15 and 16 stand rejected under 35 U.S.C. § 103(a) as being unpatentable over D'Hoore in view of Waibel ("Interactive Translation of Conversational Speech", IEEE 1996). These rejections are respectfully traversed.

Complete discussions of the Examiner's rejections are set forth in the Office Action, and are not repeated herein.

Without conceding to the propriety of the Examiner's rejection, but merely to timely advance the prosecution of the application, as the Examiner will note, independent claim 1 has been amended to more clearly define the present invention over the references relied on by the Examiner.

In particular, independent claim 1 now recites a combination of elements including "a speech modeling engine, receiving and transferring a mixed multi-lingual speech signal into a plurality of speech features; a multi-lingual baseform mapping engine, comparing a plurality of

<center>7</center>

multi-lingual query commands to obtain a plurality of multi-lingual baseforms; *a cross-lingual diphone model generation engine, coupled to the multi-lingual baseform mapping engine, selecting and combining the multi-lingual baseforms, further comprising: fixing one side contexts of the multi-lingual baseforms and mapping another side contexts of the multi-lingual baseforms to obtain a mapping result; obtaining the multi-lingual context-speech mapping data according to the mapping result;* and storing the multi-lingual context-speech mapping data in a multi-lingual model database; a speech search engine, coupled to the speech modeling engine, receiving the speech features, and locating and comparing a plurality of candidate data sets corresponding to the speech features according to the multi-lingual model database to find match probability of a plurality of candidate speech models of the candidate data sets; and a decision reaction engine, coupled to the speech search engine, selecting a plurality of resulting speech models corresponding to the speech features according to the match probability from the candidate speech models to generates a speech command." Support for this amendment may be found at least at, for example, page 10, lines 1-10 of the Specification as originally filed. Thus, no new matter has been added. Applicant respectfully submit that the combination of elements set forth in claim 1 is not disclosed or suggested by the references relied on by the Examiner.

Specifically, as set forth in amended claim 1, the cross-lingual diphone model generation engine 206 selects and combines the multi-lingual baseforms into the multi-lingual context-speech mapping data 208. The cross-lingual diphone model generation engine 206 accomplishes the selection and combination by several steps. First, the left contexts of the multi-lingual baseforms are fixed, and the right contexts of the multi-lingual baseforms are mapped to obtain a mapping result. Next, fix the right context and the left contexts are mapped to obtain the

mapping result if the right context mapping fails. Finally, multi-lingual context-speech mapping

data is obtained according to the mapping result.

However, on the contrary, referring to Col. 4, line 63-Col. 5, line 12 of D'Hoore, it is
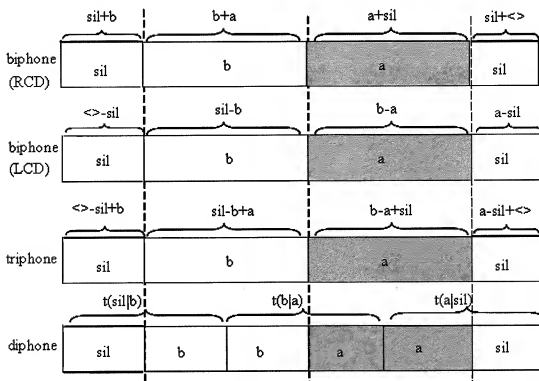
recited that:

> [t]he training procedures for single language and
> multi-language acoustic models both use standard training
> techniques; they differ in the type of data that is presented
> and the speech units that are trained. The training can be
> viewed as the construction of a database of acoustic models
> 47 covering a specific phoneme set. *The training process
> begins by training context independent models using
> Viterbi training of discrete density HMMs.* Then the
> phoneme models are automatically classified into 14
> classes. *Based on the class information, context dependent
> phoneme models are constructed.* Next, *the context
> dependent models are trained using Viterbi training of
> discrete density HMMs. The context dependent and context
> independent phoneme models are merged,* and then, *lastly,
> badly trained context dependent models are smoothed with
> the context independent models.* Such acoustic model
> training methods are well-known within the art of speech
> recognition.

(emphasis added) It is clear that D'Hoore does not disclose or suggest fixing left/right contexts

and mapping right/left contexts obtaining the multi-lingual context-speech mapping data, as set

forth in amended claim 1.

Further, on page 8 of the outstanding Office Action, the Examiner asserts that the context

dependent biphone acoustic models of D'Hoore inherently teaches the left and right context

mapping for recognition of the present invention; Applicant respectfully disagrees and submits

for something to be inherently disclosed, it cannot be just possibly disclosed, and it cannot be

just probably disclosed. Rather, what is inherently disclosed must be necessarily disclosed. In re

Oelrich, 666 F.2d 578, 581, 212 USPQ 323, 326 (CCPA 1981) and In re Rijckaert, 9 F.3d 1531, 1534, 28 USPQ2d 1955, 1957 (Fed. Cir. 1993). In this case, Applicant respectfully submits that the context dependent biphone acoustic models of D'Hoore cannot be comparable with the left and right context mapping for recognition set forth in the present invention, and thus D'Hoore does not disclose or suggest the above-mentioned feature of the present invention.

Specifically, referring to the following figure which shows a comparison between the diphone model of the present invention and the biphone model asserted by the Examiner.

	sil+b	b+a	a+sil	sil+<>
biphone (RCD)	sil	b	a	sil

	<>-sil	sil-b	b-a	a-sil
biphone (LCD)	sil	b	a	sil

	<>-sil+b	sil-b+a	b-a+sil	a-sil+<>
triphone	sil	b	a	sil

	t(sil|b)	t(b|a)	t(a|sil)	
diphone	sil	b | b	a | a	sil

In particular, compared with the diphone model of the present invention, the biphone model only depends on a single context, the right context-dependent (RCD) or the left context-dependent (LCD). For example, pronunciation of "ba" (with phoneme sequence a followed b) for LCD and RCD of biphone and diphone is shown in above figure, in which the dotted lines

10

represent the waveform boundary between phoneme "silence(sil)", "b", and "a". With respect to the LCD, waveform "b" is labeled as "sil-b" and "a" is labeled as "b-a". Based on the LCD, "b-a" can be replaced by "e-a" ("e" represents another sounded-like phoneme with the corresponding context), but cannot be able to be replaced by "b-e" because it is actually an "a" sound. With respect to the RCD, "b" is labeled as "b+a" and "a" is labeled as "a+sil". Based on the RCD, "b+a" can be replaced by "b+e", but cannot be able to be replaced by "e+a" because it is actually a "b" sound. With respect to the triphone, "b" is labeled as "sil-b+a" and "a" is labeled as "b-a+sil". Based on the triphone, "sil-b+a" can be replaced by "a-b+c", but cannot be able to be replaced by "a-e+c" because it is actually a "b" sound. In case of no corresponding pronunciation phoneme available to use, it is usually not working well based on these model.

By contrast, according to the diphone model of the present invention, t(b|a) can be replaced at least by t(e|a) and t(b|e), which provides more options in case of no corresponding pronunciation phoneme in the either side of the diphone model, so that we still could use diphone model with matched phoneme in the other side to work out moderate results. For example, in the case of using Chinese model to simulate English ASR, because there is no "v" sound in Chinese, RCD/LCD biphone and triphone will not work well for utterance with "v" sound. However, this situation would never happen to diphone model, because it could still work well to use diphone model with matched context in the other side. Another example is regarding mixed Chinese and English ASR. The RCD/LCD/triphone usually does not work as well as the diphone system for those utterance with sound "v" in cross-lingual part because there is no "v" sound in Chinese. According to the RCD/LCD/triphone, it is necessary to use English model v+*(RCD), *-v(LCD), or *-v+*(triphone) to solve the problem; however, on the contrary, the diphone model of the

11

present invention can have more choices to use Chinese or English diphone models with matched left or right context to achieve better results. It is clear that the diphone model of the present invention is distinguishable over the biphone model asserted by the Examiner. Applicant respectfully emphasizes that the present invention adopts cross-lingual diphone models to recognize the parts of the speech signal containing multiple languages and uni-lingual diphone models to recognize parts of containing only one. That is, only the parts transitioning between languages will be recognized by cross-lingual diphone models, thus avoiding the interference of different languages, which is distinguishable over the teachings of D'Hoore.

In addition, it is noted that the present invention applies a normalization method and can use models from multiple languages to implement multi-lingual recognition functions to recognize multi-lingual mixed speech signals and produce speech commands. Particularly, the present invention can be applied in a speech recognition system with a large amount of vocabulary and cross-language terms, providing significant improvement over the conventional method, which is also distinguishable over the teachings of D'Hoore.

With regard to the Examiner's reliance on Burns and Waibel, these references have only been relied on for their teachings of the dependent claims. Burns and Waibel also fail to disclose each of the above-mentioned feature set forth in claim 1.

For at least above reasons, Applicant respectfully submits that claim 1 clearly defines over the teachings of the references relied on by the Examiner. Regarding independent claim 9, it is submitted that claim 9 also clearly defines over the references relied on by the Examiner for at least the same reasons as claim 1.

PCL/QL

In addition, claims 2, 6, 7, 10 and 14-16 depend, either directly or indirectly, from independent claims 1 and 9, and are therefore allowable based on their respective dependence from independent claims 1 and 9, which are believed to be allowable.

In view of the above amendments to the claims and remarks, Applicant respectfully submits that claims 1, 2, 6, 7, 9, 10 and 14-16 clearly define the present invention over the references relied on by the Examiner. Accordingly, reconsideration and withdrawal of the rejections under 35 U.S.C. §§ 102 and 103 are respectfully requested.

Additional Claims

Claims 18-20 have been added for the Examiner's consideration. Support for new claims 18-20 may be found at least at, for example, page 9, line 24-page 10, line 10 of the Specification as originally filed and original claim 8 and 17. Thus, no new matter has been added.

Applicants respectfully submit that claims 18-20 depend, either directly or indirectly, from independent claims 1 and 9, and are therefore allowable based on their respective dependence from independent claims 1 and 9, which are believed to be allowable, as will as due to the additional novel features set forth therein.

Favorable consideration and allowance of claims 18-20 are respectfully requested.

CONCLUSION

Since the remaining patents cited by the Examiner have not been utilized to reject the claims, but merely to show the state of the prior art, no further comments are necessary with respect thereto.

PCL/QL

It is believed that a full and complete response has been made to the Office Action, and that as such, the Examiner is respectfully requested to send the application to Issue.

In the event there are any matters remaining in this application, the Examiner is invited to contact Paul C. Lewis, Registration No. 43,368, at (703) 205-8000 in the Washington, D.C. area.

If necessary, the Commissioner is hereby authorized in this, concurrent, and future replies to charge payment or credit any overpayment to Deposit Account No. 02-2448 for any additional fees required under 37.C.F.R. §§1.16 or 1.147; particularly, extension of time fees.

Dated: December 10, 2008

Respectfully submitted,

By _____
Paul C. Lewis
Registration No.: 43,368
BIRCH, STEWART, KOLASCH & BIRCH, LLP
8110 Gatehouse Road
Suite 100 East
P.O. Box 747
Falls Church, Virginia 22040-0747
(703) 205-8000
Attorney for Applicant